Biophysical Chemistry 113 (2005) 9–16

Biophysical Chemistry

www.elsevier.com/locate/bpc

Folding rate prediction using *n*-order contact distance for proteins with two- and three-state folding kinetics

Linxi Zhang[a,*], Tingting Sun[a,b]

[a] *Department of Physics, Wenzhou Normal College, Wenzhou 325027, PR China*
[b] *Department of Physics, Zhejiang University, Hangzhou 310027, PR China*

Received 4 June 2004; received in revised form 19 July 2004; accepted 21 July 2004
Available online 2 September 2004

Abstract

It is a challenging task to understand the relationship between sequences and folding rates of proteins. Previous studies are found that one of contact order (CO), long-range order (LRO), total contact distance (TCD), chain topology parameter (CTP), and effective length (L_{eff}) has a significant correlation with folding rate of proteins. In this paper, we introduce a new parameter called *n*-order contact distance (*n*OCD) and use it to predict folding rate of proteins with two- and three-state folding kinetics. A good linear correlation between the folding rate logarithm $\ln k_{\mathrm{f}}$ and *n*OCD with n=1.2, α=0.6 is found for two-state folders (correlation coefficient is −0.809, *P*-value<0.0001) and n=2.8, α=1.5 for three-state folders (correlation coefficient is −0.816, *P*-value<0.0001). However, this correlation is completely absent for three-state folders with n=1.2, α=0.6 (correlation coefficient is 0.0943, *P*-value=0.661) and for two-state folders with n=2.8, α=1.5 (correlation coefficient is −0.235, *P*-value=0.2116). We also find that the average number of contacts per residue P_m in the interval of m for two-state folders is smaller than that for three-state folders. The probability distribution $P(\gamma)$ of residue having γ pairs of contacts fits a Gaussian distribution for both two- and three-state folders. We observe that the correlations between square radius of gyration S^2 and number of residues for two- and three-state folders are both good, and the correlation coefficient is 0.908 and 0.901, and the slope of the fitting line is 1.202 and 0.795, respectively. Maybe three-state folders are more compact than two-state folders. Comparisons with *n*TCD and *n*CTP are also made, and it is found that *n*OCD is the best one in folding rate prediction.

Keywords: Folding rate; *n*-order contact distance (*n*OCD); Two- and three-state folders

1. Introduction

The folding behavior of small proteins with simple two-state kinetics and of larger proteins with three-state folding kinetics has a significant difference. The two-state folders have no visible intermediates in the course of folding, which therefore occurs as an “all-or-none” process under all experimental conditions. However, the proteins with three-state folding kinetics fold via intermediates, which accumulate during the early stages of folding when it occurs in denaturant-free water [1–4]. It has been observed that the logarithm of the folding rate of two-state folding proteins is strongly anti-correlated with a topological parameter called contact order (CO) [5], and there is complete absence of such correlation for three-state folders. Instead, the three-state folding proteins demonstrate a strong anti-correlation between their size and the folding rate logarithm [6].

Protein folding is the process by which a protein progresses from its denatured state to its specific biologically active conformation. A related important task is to understand the relationship between sequences and folding rates of proteins. Many parameters to predict folding rates have arisen. The folding rate of proteins that fold with two- or weakly three-state kinetics has a significant correlation with the average sequence separation of all contacting

[*] Corresponding author. Tel.: +86 57 187 953261; fax: +86 57 187 951328.
E-mail address: Lxzhang@hzcnc.com (L. Zhang).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.07.036

residues in the native state, defined by the parameter contact order (CO) [5]. Later, a different parameter is found to correlate better with $\ln k_f$ than CO, and the parameter is called long-range order (LRO) for a protein from the knowledge of long-range contacts in protein structure [7]. Then, Zhou and Zhou [8] brought forward a new parameter, called total contact distance (TCD), to predict folding rate of proteins. Another parameter called chain topology parameter (CTP), which is similar to contact order (CO) and is used previously to describe the complexity of the chain topology of the protein molecule, can yield much improved results [9]. We also observe a best excellent correlation between the folding rate of proteins with two-state kinetics and contact parameters (including CO, LRO, and TCD), and find that the folding rate depends on CO, LRO and TCD simultaneously [10]. Recently, Ivankov and Finkelstein [11] find that the folding rate is only correlated with the effective length of protein chains, and the correlation is somewhat simple and the interior structure of protein is considered. In this paper, we present a new parameter which contains not only the relative positions of residues but also the secondary structure of proteins. It is called n-order contact distance (nOCD), and we use this parameter to predict the folding rates of proteins with two- and three-state folding kinetics. It is shown that this parameter can correlate with the logarithm of the folding rate of two and three-state folding proteins well. We also investigate average number of contacts per residue P_m as a function of interval m ($m=|j-i|$) for two- and three-state folders. The probability distributions $P(\gamma)$ of residue having γ pairs of contacts for two- and three-state folders are considered here. At last, we discuss the square radius of gyration S^2 of proteins with two- and three-state folding kinetics. The aim is to investigate the difference of interior structure for two- and three-state folders. Some comparisons with other parameters are also made.

2. Method of calculation

2.1. Database

In this paper, we study 30 proteins with two-state kinetics and 24 proteins with three-state kinetics. The database is taken from Ref. [6]. Table 1 lists those 54 proteins, and the proteins numbered 1–30 exhibit the two-state folding within the whole range of experimental conditions and the proteins numbered 31–54 exhibit the three-state folding when the native state is much more stable than the denatured one. The data of these protein structures are taken from the Protein Data Bank (PDB) [12]. The Protein IDs used in the present study are listed in Table 1, and the references are also given. Here, n_r is the number of residues. nOCD is a new parameter and we will introduce in the next section. $\ln k_f$ is natural logarithm of the experimental folding rate (s^{-1}) measured in or extrapolated to pure water (i.e., at zero denaturant concentration).

2.2. n-order contact distance (nOCD)

It is found that the logarithms of folding rates ($\ln k_f$) of proteins that fold with two- or weakly three-state kinetics has a surprisingly simple and statistically significant correlation with a single parameter called contact order (CO) [5], and CO is defined as:

$$\mathrm{CO} = \frac{1}{n_c n_r} \sum_{|j-i|>l_{\mathrm{cut}}}^{n_c} |j-i| \tag{1}$$

In another paper [9], it is shown that $\ln k_f$ correlates well with the so-called chain topology parameter, CTP:

$$\mathrm{CTP} = \frac{1}{n_c n_r} \sum_{|j-i|>l_{\mathrm{cut}}}^{n_c} |j-i|^2 \tag{2}$$

Recently, Ivankov and Finkelstein [11] found that the folding rate mainly depends on the effective chain length (L_{eff}) and there exists the following dependence:

$$\log(k_f) \sim \mathrm{const} - L_{\mathrm{eff}}^{\mathrm{p}} \tag{3}$$

with

$$L_{\mathrm{eff}} = n_r - n_H + l_1 \times L_H \tag{4}$$

Where n_H is the number of residues in helical conformation, L_H is the number of helices, and l_1 means that we consider the whole block (a helix) as l_1 chain residues.

Because α-helices are natural candidates to the role of the internally stable and/or rapidly and independently flocks, we think that the residue in helical conformation has a different contribution to folding rate. In order to consider not only the interior structure of protein but also the secondary structure of protein, here we introduce a new parameter named n-order contact distance, nOCD:

$$n\mathrm{OCD} = \frac{1}{n_c n_r} \sum_{|j-i|>l_{\mathrm{cut}}}^{n_c} \alpha_i \alpha_j |j-i|^n \tag{5}$$

with

$$\alpha_i = \begin{cases} \alpha & \text{residue } i \quad \text{is in helical conformation} \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

where n_r is the number of amino acid residues of a protein, n_c is the number of residue–residue contacts, and i and j represent the positions of two residues in a contact. The value of α_i is related with the effect of α-helices to n-order contact distance (nOCD) for residue i. If residue i is in helical conformation, the value of α_i becomes α. Certainly, if residue i is not in helical conformation, this means there is no effect of α-helices to n-order contact distance, therefore α_i=1.0. If both residues i and j are in helical conformation, $\alpha_i\alpha_j=\alpha^2$ and if either residue i or residue j is in helical conformation, $\alpha_i\alpha_j=\alpha$ (here one of α_i and α_j is equal to 1.0). Of course, if both residues i and j are not in helical conformation, $\alpha_i\alpha_j$=1. Each residue in a protein molecule is

Table 1
The database and the values of n-order contact distance (nOCD) used in this study

No.	Protein ID	References	n_r	nOCD		$\ln k_f$
				α=0.6, n=1.2	α=1.5, n=2.8	
Proteins with two-state kinetics						
1	1APS	Van Nuland et al. [14]	98	0.797	631.9	−1.5
2	1AYE	Villegas et al. [15]	80	0.583	565.2	6.8
3	1C8C	Guerois and Serrano [16]	63	0.347	95.37	7.0
4	1C90	Perl et al. [17]	66	0.575	151.4	7.2
5	1CSP	Perl et al. [17]	67	0.558	143.2	6.5
6	1DIV	Kuhlman et al. [18]	56	0.335	81.36	6.1
7	1FKB	Main et al. [19]	107	0.660	503.9	1.5
8	1FNF_9	Plaxco et al. [20]	90	0.707	409.7	−0.9
9	1G6P	Perl et al. [17]	66	0.552	136.5	6.3
10	1HZ6	Kim et al. [21]	62	0.524	202.3	4.1
11	1IMQ	Ferguson et al. [22]	86	0.285	287.5	7.3
12	1LMB	Burton et al. [23]	80	0.174	182.7	8.5
13	1LOP	Ikura et al. [24]	164	0.632	1317	6.6
14	1MJC	Reid et al. [25]	69	0.551	150.7	5.3
15	1PGB	McCallister et al. [26]	56	0.582	186.5	6.0
16	1PNJ	Guijarro et al. [27]	86	0.698	321.0	−1.1
17	1POH	Van Nuland et al. [28]	85	0.569	495.3	2.7
18	1PSF	Schindler et al. [29]	69	0.646	285.7	3.2
19	1RIS	Otzen and Oliveberg [30]	97	0.671	563.9	5.9
20	1SHF	Plaxco et al. [31]	59	0.585	175.8	4.5
21	1SHG	Viguera et al. [32]	57	0.639	188.9	1.4
22	1SRL	Grantcharova and Baker [33]	56	0.664	200.7	4.0
23	1TEN	Clarke et al. [34]	89	0.676	371.9	1.1
24	1URN	Silow and Oliberg [35]	96	0.583	471.4	5.8
25	1WIT	Clarke et al. [36]	93	0.765	494.0	0.4
26	256B	Wittung-Stafshede et al. [37]	106	0.117	198.5	12.2
27	2ABD	Kragelund et al. [38]	86	0.276	462.1	6.6
28	2CI2	Jackson et al. [39]	64	0.607	245.1	3.9
29	2PDD	Spector and Raleigh [40]	41	0.248	45.55	9.8
30	2VIK	Choe et al. [41]	126	0.398	348.6	6.8
Proteins with three-state kinetics						
31	1A6N	Cavagnero et al. [42]	151	0.137	760.1	1.1
32	1AON	Golbik et al. [43]	155	0.550	884.8	0.8
33	1BNI	Matouschek et al. [44]	108	0.296	190.7	2.6
34	1BRS	Schreiber and Fersht [45]	89	0.405	223.1	3.4
35	1CBI	Burns et al. [46]	136	0.423	735.7	−3.2
36	1CEI	Ferguson et al. [22]	85	0.305	293.8	5.8
37	1EAL	Dalessio and Ropson [47]	127	0.431	615.3	1.3
38	1FNF_10	Cota and Clarke [48]	94	0.590	297.8	5.5
39	1HNG	Parker et al. [49]	95	0.707	552.5	1.8
40	1IFC	Burns et al. [50]	131	0.430	696.5	3.4
41	1OPA	Burns et al. [50]	133	0.479	790.3	1.4
42	1PHP_A	Parker et al. [51]	175	0.413	888.0	2.3
43	1PHP_B	Parker et al. [52]	219	0.255	966.3	−3.5
44	1QOP_A	Ogasahara and Yutani [53]	267	0.259	1815	−2.5
45	1QOP_B	Goldberg et al. [54]	396	0.299	2355	−6.9
46	1RA9	Jennings et al. [55]	159	0.590	889.6	−2.5
47	1SCE	Schymkowitz et al. [56]	101	0.408	290.7	4.2
48	1TIT	Fowler and Clarke [57]	89	0.741	441.3	3.6
49	1UBQ	Khorasanizadeh et al. [58]	76	0.553	274.0	5.9
50	2A5E	Tang et al. [59]	156	0.116	56.58	3.5
51	2CRO	Laurents et al. [60]	65	0.175	201.5	3.7
52	2LZM	Parker and Marqusee [61]	164	0.094	280.2	4.1
53	2RN2	Parker and Marqusee, [61]	155	0.438	1067	0.1
54	3CHY	Munoz et al. [62]	128	0.224	197.1	1.0

Here, n_r is number of structured residues, R_{cut}=0.80 nm (based on the C^{α} atom distance), and l_{cut}=2.

represented by C^{α} atom. Residues whose C^{α} atoms are closer than R_C are defined to form a contact, and the contact is separated by at least a residue separation cutoff value l_{cut}. Here, we choose R_C=0.80 nm and l_{cut}=2.

Using this parameter, we can obtain the values of *n*OCD and then discuss the relationships between the logarithms of folding rates (lnk_f) of proteins with two- or three-state folding kinetics and the values of *n*OCD. Through our calculation, we can get the best one.

2.3. Average number of contacts per residue P_m in the interval of m

The average number of contact per residue P_m in the interval of m is defined as:

$$P_m = \frac{N_m}{N} \tag{7}$$

here N_m is the total number of contacts with the separation in sequence between the contacting residues number i and j is equal to m, and N is the total number of residues in the proteins with two- or three-state folding kinetics. In our paper, P_m is averaged for m, m+1 and m+2. For example, P_4 is averaged for $|j-i|$=4, 5, and 6.

2.4. Probability $P(\gamma)$ of residue with forming γ pairs of contacts

In protein molecule, each residue has a different ability of forming contacts, and in general, residues in the interior of proteins have a large number of contacts. So we introduce the probability $P(\gamma)$ of residues with forming γ pairs of contacts in all residues, and it is defined as:

$$P_\gamma = \frac{N_\gamma}{N} \tag{8}$$

here N_γ is the total number of amino acid residues with forming γ pairs of contacts, and N is the total number of residues in the proteins with two- or three-state folding kinetics, respectively.

In the meantime, we can also calculate the square radius of gyration S^2 from the 3D coordinates of the residues, and can know the average dimensions of two- or three-state folders.

3. Results and discussion

3.1. Correlation between the experimental observed lnk_f and n-order contact distance (nOCD) for two- and three-state folders

According to Eq. (5), we first calculate the values of *n*-order contact distance (*n*OCD) with different values of *n* and α for two- and three-state folders. Then, we can get the correlations between the experimental observed lnk_f and *n*-order contact distance (*n*OCD). The absolute value of correlation coefficient between lnk_f and *n*OCD for two-state folders with n=1.2 and α=0.6 is the largest one. In the meantime, the absolute value of correlation coefficient between lnk_f and *n*OCD for three-state folders with n=2.8, and α=1.5 is the largest one. In fact, a large absolute value of correlation coefficient means a good linear fit. If the absolute value of correlation coefficient equals to 1, the plot has the best linear fit. In Fig. 1, we plot the correlations between lnk_f and *n*OCD with n=1.2, α=0.6 (Fig. 1a) and n=2.8, α=1.5 (Fig. 1b) for two- and three-state folders, respectively. We can see from Fig. 1(a) that *n*OCD with n=1.2 and α=0.6 has a correlation coefficient of −0.809 (P-value<0.0001) with the logarithms of folding rates for the two-state folders, while for three-state folders, *n*OCD with n=1.2 and α=0.6 has a correlation coefficient of 0.0943 (P-value=0.661) with lnk_f. The line in Fig. 1(a) represents the

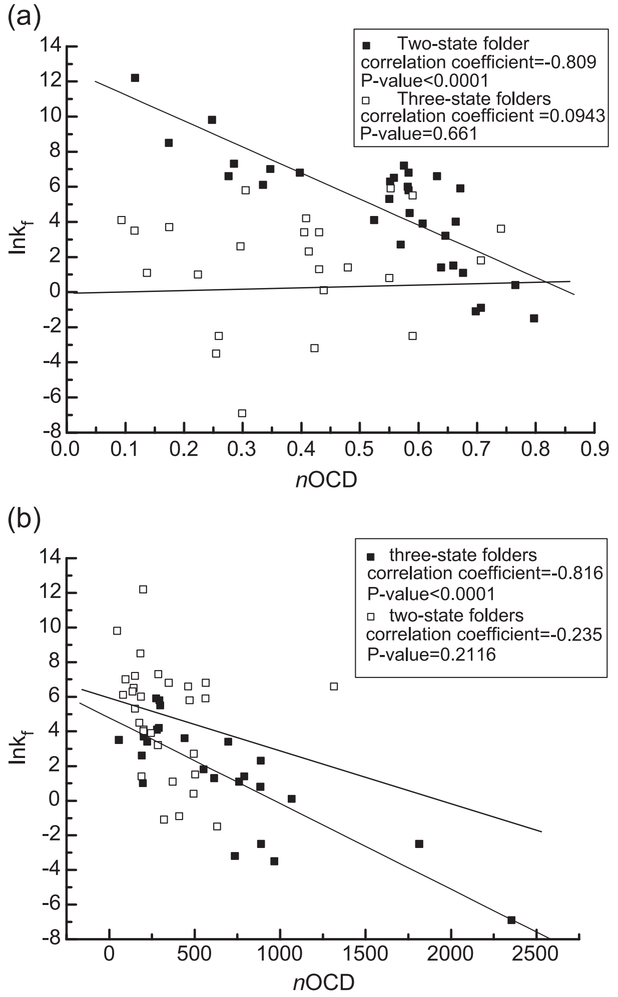

Fig. 1. Correlation between the experimental observed lnk_f and n-order contact distance (nOCD) for two- and three-state folders. Here, (a) two- and three-state folders with n=1.2, andα=0.6 (correlation coefficient=−0.809, P-value<0.0001 for two-state folders (■), and correlation coefficient=0.0943, and P-value=0.661 for three-statefolders (□)); (b) two- and three-state folders with n=2.8, and α=1.5 (correlation coefficient=−0.816, and P-value<0.0001 for three-state folders (■), and correlation coefficient=−0.235, and P-value=0.2116 for two-state folders (□)).

best linear fit for two-state folders in our calculation. The P-value is associated with the correlation coefficient, and low P-value suggests that the observed correlation is highly unlike to have arisen by chance. There is a strong inverse relationship between nOCD and lnk_f for two-state folders with n=1.2 and α=0.6.

We also observe that the value of nOCD with n=2.8 and α=1.5 shows a good relationship with folding rate for the three-state folders. The correlation coefficient is as significant as −0.816 and the P-value is less than 0.0001. However, for two-state folders, the relationship between nOCD and lnk_f is not good. In this case, the correlation coefficient is −0.235, and the P-value is 0.2116. Therefore, we can conclude that there exists a strong negative correlation between the nOCD with n=2.8, α=1.5 and the logarithm of folding rates for three-state folders. If the outlying point (nOCD=2355 for 1QOP_B) is not considered in Fig. 1(b), the correlation coefficient becomes −0.748 for 23 three-state folders.

Fig. 2 gives the correlation coefficient for lnk_f as a function of nOCD with different values of n, α, and R_C. In Fig. 2(a), we draw some plots of correlation coefficient versus α with different n and R_C for two-state folders, and can find that when n=1.2 and α=0.6, the absolute value of correlation coefficient is the largest one for two-state folders (=0.809). This means that in this point, nOCD has the best linear relationship with lnk_f for two-state folders. In the same way, we also plot the curves of correlation coefficient versus α with different n and R_C for three-state folders. It is investigated that when n=2.8 and α=1.5, the absolute value of correlation coefficient is the largest one for three-state folders. The value is 0.816. In this figure, the point of α=1 and n=1 means the correlation coefficient for lnk_f~CO, in the meantime, the point of α=1 and n=2 shows CTP. From these comparisons, we find the absolute values of correlation coefficients for lnk_f~CO and CTP are both smaller than nOCD. Therefore, we can conclude that nOCD is better than CO and/or CTP.

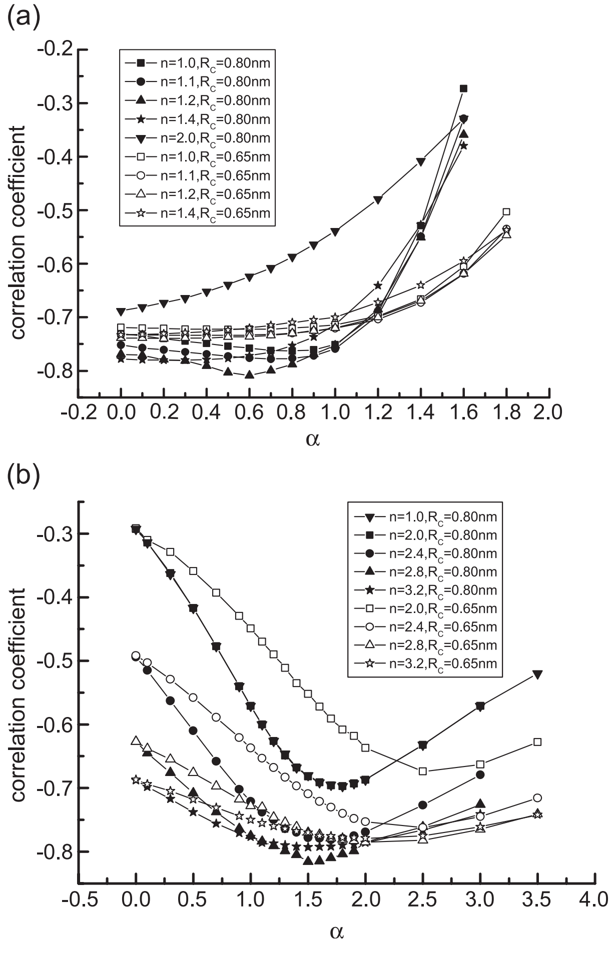

Fig. 2. Correlation coefficient for lnk_f~nOCD with different values of α, n and R_C. Here, (a) two-state folders, and (b) three-state folders.

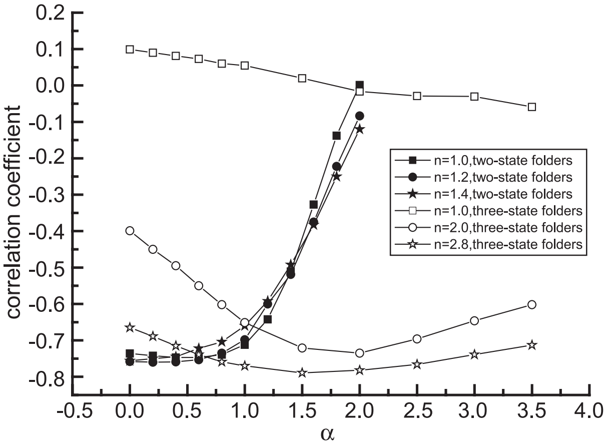

Fig. 3. Correlation coefficient for lnk_f~nTCD with different values of α and n for two- and three-state folders.

Similar to the definition of nOCD, here we also present another parameter, nTCD, which is defined as:

$$nTCD = \frac{1}{n_{\mathrm{r}}^2} \sum_{|j-i|>l_{\mathrm{cut}}}^{n_{\mathrm{c}}} \alpha_i \alpha_j |j-i|^n \qquad (9)$$

If n=1 and α=1, nTCD becomes total contact distance (TCD) [6]. According to this definition, we calculate the correlation coefficient for lnk_f as a function of nTCD with different n and α for two- and three-state folders, and the results are given in Fig. 3. In Fig. 3, we find that nTCD of the two-state folders has the best linear relationship with lnk_f for n=1.2 and α=0.4, and the absolute value of correlation coefficient is 0.759. While for three-state folders, the largest absolute value of correlation coefficient is at the point of n=2.8, α=1.5, and the value is 0.789. However, they are both smaller than that of nOCD. In Fig. 3, nOCD with n=1 and α=1 represents TCD. The absolute value of correlation coefficient for TCD is also smaller than nOCD. Therefore, we regard the nOCD as the better parameter to predict folding rate. In fact, comparisons with Figs. 1 and 2, we can also find that the values of n in both

nOCD and nTCD for two-state folders are all less than those for three-state folders, and the values of α in both nOCD and nTCD for two-state folders are also less than those for three-state folders. Similar results are obtained for nOCD and nTCD.

3.2. Average number of contacts per residue P_m in the interval of m for two- and three-state folders

In order to give the reason why there exists different relationship between lnk_f and nOCD for proteins with two- and three-state folding kinetics, here we investigate the differences of interior structures for two- and three-state folders. According to Eq. (7), we calculate the average number of contacts per residue P_m in the interval of m for two- and three-state folders, and the results are shown in Fig. 4. It is shown that at the same m, the value of P_m for two-state folders is smaller than that for three-state folders. In other words, the average number of contacts per residue for three-state folders is larger than that for two-state folders. For example, in Fig. 4, P_m with m=40 for three-state folders is 32.1% larger than that for two-state folders. Meanwhile, P_m with m=10 for three-state folders is 116.7% larger than that for two-state folders. Maybe three-state folders are more compact than two-state folders.

We also find that there were no contacts with the separation in sequence between the residues which is larger than 163 for the two-state folders. The contacts do exist for three-state folders with m>163. However, there are no contacts any more with m>307 for three-state folders.

3.3. The probability distribution P(γ) of residue having γ pairs of contacts for two- and three-state folders

We use Eq. (8) to calculate the probability of amino acid residues with forming γ pairs of contacts in all residues for two- and three-state folders, and the results are shown in Fig. 5. In Fig. 5, we find that the probability distributions $P(\gamma)$ for two- and three-state folders both fit

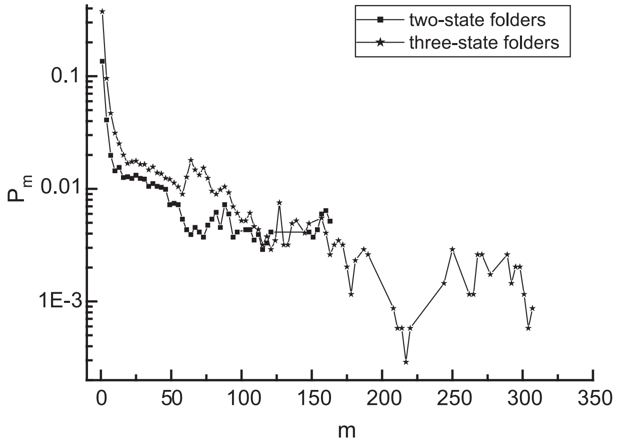

Fig. 4. Average number of contacts per residue P_m in the interval of m as a function of m (m=|$j-i$|) for two- and three-state folders. Here, i and j represent residues i and j, and P_m is also averaged for m, m+1, and m+2. For example, P_4 is averaged for |$j-i$|=4, 5, and 6.

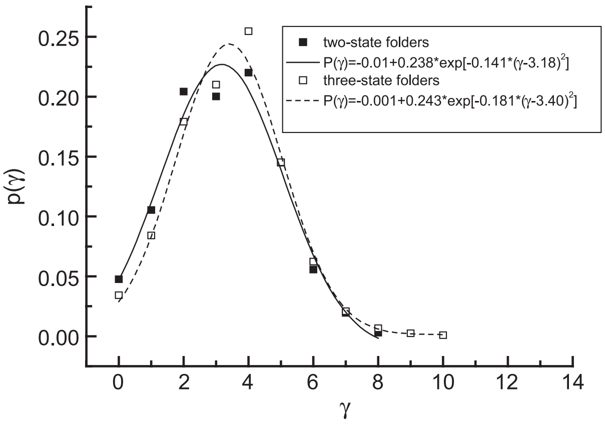

Fig. 5. The probability distribution $P(\gamma)$ of residues having γ pairs of contacts as a function of γ for two- and three-state folders. The solid line is Gaussian distribution for two-state folders and the dash line is for three-state folders.

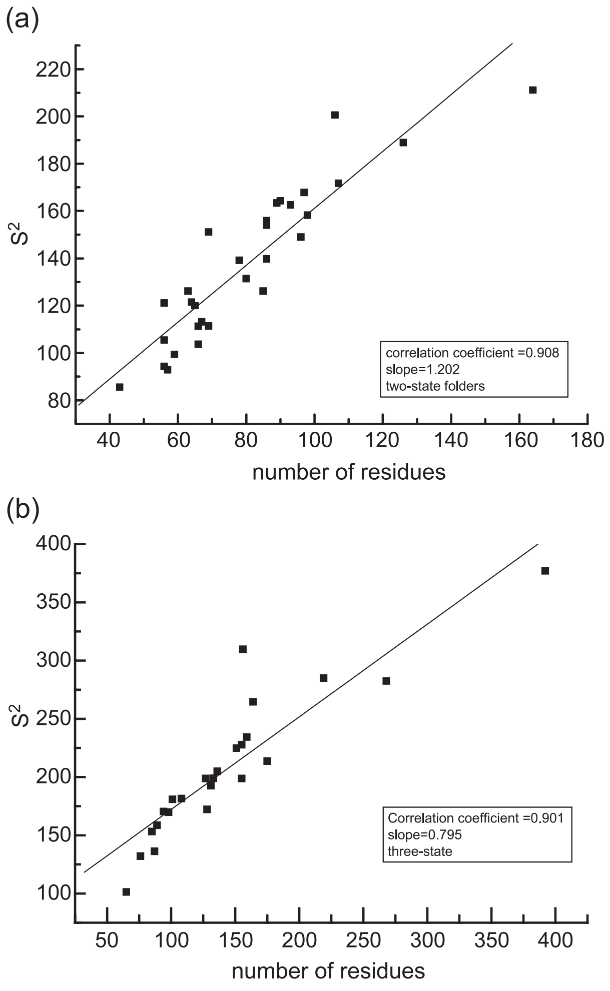

Fig. 6. Square radius of gyration S^2 as a function of number of residues for two- and three-state folders. Here, (a) two-state folders, and (b) three-state folders.

Gaussian distribution. The probability distribution $P(\gamma)$ of amino acid residues with forming γ pairs of contacts in all residues can be expressed as Gaussian distribution [13]:

$$P(\gamma) = P_0 + a\exp\left[-b(\gamma - \gamma_c)^2\right] \tag{10}$$

with P_0=−0.01, a=0.238, b=0.141 and γ_c=3.18 for two-state folders and P_0=−0.001, a=0.243, b=0.181 and γ_c=3.40 for three-state folders. Here, there are no any error bars in Figs. 4 and 5 because the total number of amino acid residues with forming γ pairs of contacts N_γ is very exact in our calculation.

It is shown that the probability of amino acid residues $P(\gamma)$ with forming γ pairs of contacts in all residues for two-state folders is larger than that for three-state folders when γ=0, 1 and 2. When $\gamma \geq 3$, the value of $P(\gamma)$ for three-state folders is larger than that for two-state folders. We also find that when γ=9 and 10, the probability $P(\gamma)$ for three-state folders is not equal to zero, while the probability $P(\gamma)$ for two-state folders equals to zero.

3.4. The square radius of gyration S^2 for two- and three-state folders

The square radius of gyration S^2 of polymer chain is meaningful in the research of polymer chain, and it is also a useful parameter to assess the size of non-spherical polymer chain. So we use S^2 to assess the shapes of two- and three-state folders. Here, we compute the square radius of gyration S^2, and plot square radius of gyration S^2 as a function of total number of residues for two- and three-state folders in Fig. 6. This measure is sensitive to the mass distribution of proteins. The fit lines of two- and three-state folders are given in Fig. 6(a) and (b), respectively. For two-state folders, the correlation coefficient is 0.908 and for three-state folders, the correlation coefficient is 0.901. There are both linear correlations between square radius of gyration S^2 and number of residues for two- and three-state folders. The slope is 1.202 and 0.795 for two- and three-state folders, respectively. It suggests that the proteins with three-state folding kinetics are more compact than the proteins with two-state folding kinetics. This investigation may provide some insights into the folding rate of proteins.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (nos. 20174036, 20274040), and the Natural Science Foundation of Zhejiang Province (No. 10102). We also thank the referees for their critical reading of the manuscript and their good ideas.

References

[1] S.E. Jackson, How do small single-domain proteins fold? Fold. Des. 3 (1998) R81–R91.

[2] A.R. Fersht, Kinetics of protein folding, in: G.L. Hadler (Ed.), Structure and Mechanism in Protein Science, W.H. Freeman and Co, New York, 1999, pp. 540–572.

[3] A.R. Fersht, Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 1525–1529.

[4] W.A. Eaton, V. Munoz, S.J. Hagen, G.S. Jas, L.J. Lapidus, S.R. Henry, J. Hofrichter, Fast kinetics and mechanisms in protein folding, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 327–359.

[5] K.W. Plaxco, K.T. Simons, D. Baker, Contact order, transition state placement and the refolding rates of single domain proteins, J. Mol. Biol. 277 (1998) 985–994.

[6] O.V. Galzitskaya, S.O. Garbuzynskiy, D.N. Ivankov, A.V. Finkelstein, Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics, Proteins 51 (2003) 162–166.

[7] M.M. Gromiha, S. Selvaraj, Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long range order to folding rate prediction, J. Mol. Biol. 310 (2001) 27–32.

[8] H.Y. Zhou, Y.Q. Zhou, Folding rate prediction using total contact distance, Biophys. J. 82 (2002) 458–463.

[9] B. Nolting, W. Schalike, P. Hampel, F. Grundig, S. Gantert, N. Sips, W. Bandlow, P.X. Qi, Structural determinants of the rate of protein folding, J. Theor. Biol. 223 (2003) 299–307.

[10] L.X. Zhang, J. Lin, Z.T. Jiang, A.G. Xia, Folding rate prediction based on neural network model, Polymer 44 (2003) 1751–1757.

[11] D.N. Ivankov, A.V. Finkelstein, Prediction of protein folding rates from the amino acid sequence-predicted secondary structure, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 8942–8944.

[12] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, et al., The Protein Data Bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.

[13] T.T. Sun, L.X. Zhang, J. Chen, Analysis of structural statistical properties of proteases and nonproteases, Polymer 45 (2004) 1045–1053.

[14] N.A. Van Nuland, F. Chiti, N. Taddei, G. Raugei, G. Ramponi, C.M. Dobson, Slow folding of muscle acylphosphatase in the absence of intermediates, J. Mol. Biol. 283 (1998) 883–891.

[15] V. Villegas, A. Azuaga, L. Catasus, D. Reverter, P.L. Mateo, F.X. Aviles, L. Serrano, Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2, Biochemistry 34 (1995) 15105–15110.

[16] R. Guerois, L. Serrano, The SH3-fold family: experimental evidence and prediction of variations in the folding pathways, J. Mol. Biol. 304 (2000) 967–982.

[17] D. Perl, C. Welker, T. Schindler, K. Schroder, M.A. Marahiel, R. Jaenicke, F.X. Schmid, Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins, Nat. Struct. Biol. 5 (1998) 229–235.

[18] B. Kuhlman, D.L. Luisi, P.A. Evans, D.P. Raleigh, Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9, J. Mol. Biol. 284 (1998) 1661–1670.

[19] E.R. Main, K.F. Fulton, S.E. Jackson, Folding pathway of FKBP12 and characterisation of the transition state, J. Mol. Biol. 291 (1999) 429–444.

[20] K.W. Plaxco, C. Spitzfaden, I.D. Campbell, C.M. Dobson, A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules, J. Mol. Biol. 270 (1997) 763–770.

[21] D.E. Kim, C. Fisher, D. Baker, A breakdown of symmetry in the folding transition state of protein L, J. Mol. Biol. 298 (2000) 971–984.

[22] N. Ferguson, A.P. Capaldi, R. James, C. Kleanthous, S.E. Radford, Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9, J. Mol. Biol. 286 (1999) 1597–1608.

[23] R.E. Burton, G.S. Huang, M.A. Daugherty, P.W. Fullbright, T.G. Oas, Microsecond protein folding through a compact transition state, J. Mol. Biol. 263 (1996) 311–322.

[24] T. Ikura, T. Hayano, N. Takahashi, K. Kuwajima, Fast folding of *Escherichia coli* cyclophilin A: a hypothesis of a unique hydrophobic core with a phenylalanine cluster, J. Mol. Biol. 297 (2000) 791–802.

[25] K.L. Reid, H.M. Rodriguez, B.J. Hillier, L.M. Gregoret, Stability and folding properties of a model beta-sheet protein, *Escherichia coli* CspA, Protein Sci. 7 (1998) 470–479.

[26] E.L. McCallister, E. Alm, D. Baker, Critical role of beta-hairpin formation in protein G folding, Nat. Struct. Biol. 7 (2000) 669–673.

[27] J.I. Guijarro, C.J. Morton, K.W. Plaxco, I.D. Campbell, C.M. Dobson, Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy, J. Mol. Biol. 276 (1998) 657–667.

[28] N.A. Van Nuland, W. Meijberg, J. Warner, V. Forge, R.M. Scheek, G.T. Robillard, C.M. Dobson, Slow cooperative folding of a small globular protein HPr, Biochemistry 37 (1998) 622–637.

[29] T. Schindler, M. Herrler, M.A. Marahiel, F.X. Schmid, Extremely rapid protein folding in the absence of intermediates, Nat. Struct. Biol. 2 (1995) 663–673.

[30] D.E. Otzen, M. Oliveberg, Salt-induced detour through compact regions of the protein folding landscape, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 11746–11751.

[31] K.W. Plaxco, J.I. Guijarro, C.J. Morton, M. Pitkeathly, I.D. Campbell, C.M. Dobson, The folding kinetics and thermodynamics of the Fyn-SH3 domain, Biochemistry 37 (1998) 2529–2537.

[32] A.R. Viguera, L. Serrano, M. Wilmanns, Different folding transition states may result in the same native structure, Nat. Struct. Biol. 3 (1996) 874–880.

[33] V.P. Grantcharova, D. Baker, Folding dynamics of the src SH3 domain, Biochemistry 36 (1997) 15685–15692.

[34] J. Clarke, S.J. Hamill, C.M. Johnson, Folding and stability of afibronectin type III domain of human tenascin, J. Mol. Biol. 270 (1997) 771–778.

[35] M. Silow, M. Oliverberg, High-energy channeling in protein folding, Biochemistry 36 (1997) 7633–7637.

[36] J. Clarke, E. Cota, S.B. Fowler, S.J. Hamill, Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway, Struct. Fold. Des. 7 (1999) 1145–1153.

[37] P. Wittung-Stafshede, J.C. Lee, J.R. Winkler, H.B. Gray, Cytochrome b562 folding triggered by electron transfer: approaching the speed limit for formation of a four-helix-bundle protein, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 6587–6590.

[38] B.B. Kragelund, C.V. Robinson, J. Knudsen, C.M. Dobson, F.M. Poulsen, Folding of a four-helix bundle: studies of acyl-coenzyme A binding protein, Biochemistry 34 (1995) 7217–7224.

[39] S.E. Jackson, A.R. Fersht, Folding of chymotrypsin inhibitor 2: 1. Evidence for a two-state transition, Biochemistry 30 (1991) 10428–10435.

[40] S. Spector, D.P. Raleigh, Submillisecond folding of the peripheral subunit-binding domain, J. Mol. Biol. 293 (1999) 763–768.

[41] S.E. Choe, P.T. Matsudaira, J. Osterhout, G. Wagner, E.I. Shakhnovich, Folding kinetics of villin 14T, a protein domain with a central beta-sheet and two hydrophobic cores, Biochemistry 37 (1998) 14508–14518.

[42] S. Cavagnero, H.J. Dyson, P.E. Wright, Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin, J. Mol. Biol. 285 (1999) 269–282.

[43] R. Golbik, R. Zahn, S.E. Harding, A.R. Fersht, Thermodynamic stability and folding of GroEL minichaperones, J. Mol. Biol. 276 (1998) 505–515.

[44] A. Matouschek, J.T. Kellis Jr., L. Serrano, M. Bycroft, A.R. Fersht, Transient folding intermediates characterized by protein engineering, Nature 346 (1990) 440–445.

[45] G. Schreiber, A.R. Fersht, The refolding of cis- and transpeptidyl-prolyl isomers of barstar, Biochemistry 32 (1993) 11195–11203.

[46] L.L. Burns, P.M. Dalessio, I.J. Ropson, Folding mechanism of three structurally similar beta-sheet proteins, Proteins 33 (1998) 107–118.

[47] P.M. Dalessio, I.J. Ropson, Beta-sheet proteins with nearly identical structures have different folding intermediates, Biochemistry 39 (2000) 860–871.

[48] E. Cota, J. Clarke, Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module, Protein Sci. 9 (2000) 112–120.

[49] M.J. Parker, C.E. Dempsey, M. Lorch, A.R. Clarke, Acquisition of native beta-strand topology during the rapid collapse phase of protein folding, Biochemistry 36 (1997) 13396–13405.

[50] L.L. Burns, P.M. Dalessio, I.J. Ropson, Folding mechanism of three structurally similar beta-sheet proteins, Proteins 33 (1998) 107–118.

[51] M.J. Parker, J. Spencer, A.R. Clarke, An integrated kinetic analysis of intermediates and transition states in protein folding reactions, J. Mol. Biol. 253 (1995) 771–786.

[52] M.J. Parker, R.B. Sessions, I.G. Badcoe, A.R. Clarke, The development of tertiary interactions during the folding of a large protein, Fold. Des. 1 (1996) 145–156.

[53] K. Ogasahara, K. Yutani, Unfolding–refolding kinetics of the tryptophan synthase alpha subunit by CD and fluorescence measurements, J. Mol. Biol. 236 (1994) 1227–1240.

[54] M.E. Goldberg, G.V. Semisotnov, B. Friguet, K. Kuwajima, O.B. Ptitsyn, S. Sugai, An early immunoreactive folding intermediate of the tryptophan synthease beta 2 subunit is a "molten globule", FEBS Lett. 263 (1990) 51–56.

[55] P.A. Jennings, B.E. Finn, B.E. Jones, C.R. Matthews, A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli*: verification and refinement of a four-channel model, Biochemistry 32 (1993) 3783–3789.

[56] J.W. Schymkowitz, F. Rousseau, L.R. Irvine, L.S. Itzhaki, The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping, Struct. Fold. Des. 8 (2000) 89–100.

[57] S.B. Fowler, J. Clarke, Mapping the folding pathway of an immunoglobulin domain. Structural detail from phi value analysis and movement of the transition state, Structure 9 (2001) 355–366.

[58] S. Khorasanizadeh, I.D. Peters, H. Roder, Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues, Nat. Struct. Biol. 3 (1996) 193–205.

[59] K.S. Tang, B.J. Guralnick, W.K. Wang, A.R. Fersht, L.S. Itzhaki, Stability and folding of the tumour suppressor protein p16, J. Mol. Biol. 285 (1999) 1869–1886.

[60] D.V. Laurents, S. Corrales, M. Elias-Arnanz, P. Sevilla, M. Rico, S. Padmanabhan, Folding kinetics of phage 434 Cro protein, Biochemistry 39 (2000) 13963–13973.

[61] M.J. Parker, S. Marqusee, The cooperativity of burst phase reactions explored, J. Mol. Biol. 293 (1999) 1195–1210.

[62] V. Munoz, E.M. Lopez, M. Jager, L. Serrano, Kinetic characterization of the chemotactic protein from *Escherichia coli*, Che Y. Kinetic analysis of the inverse hydrophobic effect, Biochemistry 33 (1994) 5858–5866.